A penalized algorithm for event-specific rate models for recurrent events

Olivier Bouaziz, Agathe Guilloux

Biostatistics (2014), **0**, 0, *pp.* 1–20
doi:10.1093/biostatistics/revision2_agathe

A penalized algorithm for event-specific rate models for recurrent events

O. Bouaziz

MAP5, UMR CNRS 8145

University Paris Descartes,

Paris, France,

olivier.bouaziz@parisdescartes.fr

and

A. Guilloux

LSTA, University Pierre et Marie Curie,

4 place Jussieu,

Paris, France,

agathe.guilloux@upmc.fr

Summary

We introduce a covariate-specific total variation penalty in two semiparametric models for the rate function of recurrent event process. The two models are a stratified Cox model, introduced in Prentice et al. (1981), and a stratified Aalen's additive model. We show the consistency and asymptotic normality of our penalized estimators. We demonstrate, through a simulation study, that our estimators outperform classical estimators for small to moderate sample sizes. Finally an application to the bladder tumour data of Byar (1980) is presented.

Key words: Recurrent events process; total variation penalization; Aalen model; Cox model.

1. Introduction

Recurrent events arise in clinical or epidemiological studies when each subject experiences repeated events over time. Clinical examples include repetition of asthma attacks, epileptic seizures or tumour recurrences for individual patients. In this context, proportional hazards models have been largely studied in the literature to model the rate or mean functions of recurrent event data. For instance, Andersen & Gill (1982) introduce a conditional Cox model where the recurrent events process is assumed to be a Poisson process. Similar proportional hazards models and extensions are considered in Lawless & Nadeau (1995), Lin et al. (1998), Lin et al. (2000) and Cai & Schaubel (2004).

To model rate functions in a recurrent events context, a different approach consists in fitting a Cox model for each event count. Along these lines, Prentice et al. (1981) introduce two stratified proportional hazards models with event-specific baseline hazards and regression coefficients. Gap times and conditional models are presented in their paper and a marginal event-specific model is studied in Wei et al. (1989). We refer to Kelly & Lim (2000) for a complete review of existing Cox-based recurrent event models.

Additive models provide an useful alternative to proportional hazards models. For classical counting processes, the Aalen model was first introduced in Aalen (1980) and is extensively studied in McKeague (1988), Huffer & McKeague (1991), Lin & Ying (1994). It is considered in the context of recurrent events in Scheike (2002). We propose in this paper to consider an event-stratified version of the Aalen model, in the manner of Prentice et al. (1981).

As demonstrated in what follows, event-stratified models allow more flexibility but suffer from over-parametrization when the sample size is not large enough compared to the numbers of covariates and recurrent events. To address this drawback, we construct estimators which do not vary much between two consecutive recurrent events, or equivalently with small total variations (see Section 2.4 for details). To achieve this goal, we consider minimizers of empirical risks penalized via a covariate-specific total variation penalty. As such, our algorithms are part of the class of fused lassos algorithms. The latter have been introduced and studied, for noised piecewise constant signals, by Tibshirani et al. (2005), Rinaldo (2009) or Harchaoui & Lévy-Leduc (2010). Lasso estimators in the context of survival analysis with high dimensional covariates have been introduced and studied in Tibshirani (1997), Huang et al. (2013) in the Cox model and Martinussen & Scheike (2009b), Gaiffas & Guilloux (2012) in the Aalen model, among others.

The settings of recurrent events with a terminal event, and the two models studied in this paper are presented in the next section. In paragraph 2.4, we describe our novel algorithm, which involves a total variation penalization of criteria, specific to either the multiplicative or additive models. It requires preliminary details on inference in these two models, which are given in paragraph 2.2 and 2.3. Consistency and asymptotics distributions of our estimators are derived in Section 3. Simulation studies and a real data analysis are provided in Sections 4 and 5. We conclude with a discussion in Section 6.

2. Models and algorithm

2.1 *Models*

Let D denote the time of the terminal event and $\tilde{N}(t)$ the càdlàg process that counts the number of recurrent events occurring in the interval $(0, t]$, with the convention $\tilde{N}(0) = 0$. As no recurrent events can occur after D, the process $\tilde{N}$ has jumps only on $(0, D]$. The p-dimensional process of covariates is denoted by X and is assumed to be left continuous. The event-specific rate function of the process $\tilde{N}$, denoted by ρ_0, is defined as

$$\mathbb{E}\big[d\tilde{N}(t) \mid X(t), D \geqslant t, \tilde{N}(t-) = s-1\big]\mathbf{1}(\tilde{N}(t-) = s-1) = \mathbf{1}(D \geqslant t, \tilde{N}(t-) = s-1)\rho_0(t, s, X(t))dt, \tag{2.1}$$

for $s = 1, 2, \ldots$ and t in $A_s = \{t : \mathbb{P}[\tilde{N}(t-) = s-1, D \geqslant t] > 0\}$ and is null outside of A_s. The set A_s represents the time intervals where the s-th recurrent events can occur with a positive probability and lies in $\mathrm{Supp}(D)$, the support of D. Apart from the stratification, this definition of the rate function can be found in Scheike (2002).

We consider two semiparametric models for the function ρ_0. The first one is an event-specific multiplicative rate model introduced in Prentice et al. (1981). In this model, the rate function is specified, for $s = 1, 2, \ldots$ and t in A_s, by

$$\rho_0(t, s, X(t)) = \alpha_0(t, s) \exp\left(X(t)\beta_0(s)\right) \tag{2.2}$$

where for each event number s, $\beta_0(s)$ is an unknown p-dimensional vector of parameters and α_0 is an unknown baseline function.

Following Scheike (2002), and Zeng & Cai (2010), we also propose to consider its additive counterpart. The rate function in our event-specific additive model is then for $s = 1, 2, \ldots$ and t in A_s:

$$\rho_0(t, s, X(t)) = \alpha_0(t, s) + X(t)\beta_0(s). \tag{2.3}$$

When β_0 is assumed to be constant over the events, models (2.2) and (2.3) are usually referred to as stratified Cox and Aalen models (see e.g. Martinussen & Scheike (2006), page 190). To insist on the constancy of β_0 (as a function of s), these particular cases of models are hereafter designated as "constant coefficient models".

As, in practice, the individuals experience only a finite number of recurrent events, we will concentrate on the estimation of the rate function for the first B events, where B is an user-chosen integer (see the example in Section 5). Mathematically, this means that we only consider the observation of the process $\tilde{N}$ on the interval $[0, E(B)]$, where $E(B)$ is the hitting time of $[B, \infty)$. Equivalently, we consider that we observe the stopped process N^*, defined through $N^*(t) = \tilde{N}(t \wedge E(B))$, for all $t \geqslant 0$. Noticing that, for all $s = 1, \ldots, B$ and all $t \geqslant 0$, $\{N^*(t-) = s - 1\} = \{\tilde{N}(t-) = s - 1\} \subset \{t \leqslant E(B)\}$, the event-specific rate function of N^* equals the one of $\tilde{N}$, such that equation (2.1) holds with $\tilde{N}$ replaced by N^*. This is equivalent to assuming that the total number of observed events is almost surely bounded by B, which is the classical framework for inference for recurrent event processes (see e.g. Scheike (2002), Dauxois & Sencey (2009) or Bouaziz et al. (2013)).

We consider the problem of estimating the unknown parameter $\beta_0 = (\beta_0(1), \ldots, \beta_0(B))$, in stratified models (2.2) and (2.3) on the basis of data from n independent individuals. Introducing the censoring time C, the data consist of n independent replications $\{N_i(t), T_i, \delta_i, X_i(t), t \leqslant \tau\}$, $i = 1, \ldots, n$, where $N_i(t) = N_i^*(t \wedge C_i)$, $T_i = D_i \wedge C_i$ is the minimum between D_i and C_i, $\delta_i = \mathbf{1}(D_i \leqslant C_i)$, $(X_i(t), 0 \leqslant t \leqslant T_i)$ is the covariates process and τ represents the end-point of the study. In addition, we define the event-specific at-risk function Y^s and the overall at-risk function Y. For each individual i, for all t in $[0, \tau]$:

$$Y_i^s(t) = \mathbf{1}(T_i \geqslant t, N_i(t-) = s-1), \quad Y_i(t) = \sum_{s=1}^{B} Y_i^s(t) = \mathbf{1}(T_i \geqslant t).$$

Let $A_s^\tau = A_s \cap [0, \tau]$. The following two assumptions are mandatory to perform estimation.

Assumption 1 For all $s = 1, \ldots, B$, and t in A_s^τ, $\mathbb{E}[Y^s(t)] > 0$ and $\mathbb{P}[E(B) \leqslant \tau] > 0$.

The fist part is classical in survival analysis (see for instance Andersen et al. (1993)). The second part implies that, for $s = 1, 2, \ldots, B$, the sets A_s^τ are non-empty. Note also that the processes Y_i^s are almost surely null on the complementary of A_s in $[0, \tau]$.

Assumption 2 For all $s = 1, \ldots, B$, and t in A_s^τ,

$$\mathbb{E}\big[dN^*(t) \mid X(t), D \wedge C \geqslant t, N^*(t-) = s-1\big] = \mathbb{E}\big[dN^*(t) \mid X(t), D \geqslant t, N^*(t-) = s-1\big].$$

This assumption is classical in recurrent events context and can be found for instance in Lin et al. (2000). It is the analog of the independent right censoring definition III.2.1. of Andersen et al. (1993). We refer the reader to the Supplementary Material for a discussion on these assumptions. In particular, sufficient conditions are presented for these assumptions to hold.

In our framework, the unknown vector of parameters β_0 has $p \times B$ unknown coefficients to be estimated. For reasonable sizes of sample n, these models are over-parametrized in the sense that, when $\sqrt{n} \leqslant p \times B$, the estimators show very poor behaviour (see Section 4 for an illustration). On the other hand, simpler forms of models (2.2) and (2.3), in which the unknown parameter does not change with the event, $\beta_0(s) = \beta_0$, might be too poor to accurately fit the data (see also Section 4 and the discussion in Kelly & Lim (2000)). In this paper, we aim at providing estimators realizing a compromise between these two situations.

2.2 *Inference in the multiplicative model*

As in Prentice et al. (1981), in the multiplicative event-specific model (2.2), an estimator $\hat{\beta}_{ES/mult}$ of the unknown parameter $\beta_0 \in \mathbb{R}^{p \times B}$ is defined as the maximizer of the stratified partial log-likelihood, or equivalently as

$$\begin{aligned}\hat{\beta}_{ES/mult} &\in \underset{\beta \in \mathbb{R}^{p \times B}}{\operatorname{argmin}} L_n^{PL}(\beta) \\ &= \underset{\beta \in \mathbb{R}^{p \times B}}{\operatorname{argmin}} \left[-\frac{1}{n} \sum_{s=1}^{B} \sum_{i=1}^{n} \int_{[0,\tau]} \left\{ X_i(t)\beta(s) - \log\left(\sum_{j=1}^{n} Y_j^s(t) \exp\left(X_j(t)\beta(s)\right) \right) \right\} Y_i^s(t) dN_i(t) \right].\end{aligned} \tag{2.4}$$

An estimator $\hat{\beta}_{C/mult}$ in the constant coefficient model is defined as

$$\hat{\beta}_{C/mult} \in \underset{\beta \in \mathbb{R}^{p}}{\operatorname{argmin}} \left[-\frac{1}{n} \sum_{i=1}^{n} \sum_{s=1}^{B} \int_{[0,\tau]} \left\{ X_i(t)\beta - \log\left(\sum_{j=1}^{n} Y_j^s(t) \exp\left(X_j(t)\beta\right) \right) \right\} Y_i^s(t) dN_i(t) \right]. \tag{2.5}$$

The right term of equation (2.5) is the standard log-likelihood of a stratified Cox model (see for instance Therneau (2000) p.44-45 or Kalbfleisch & Prentice (2011) p.118-119).

2.3 *Inference in the additive model*

As noticed in Martinussen & Scheike (2009a,b) or Gaiffas & Guilloux (2012), in the usual additive hazards model, the estimator $\hat{\beta}_{ES/add}$ of the unknown parameter $\beta_0 \in \mathbb{R}^{p \times B}$ can be written as the minimizer of a (partial) least-squares criterion:

$$\hat{\beta}_{ES/add} \in \underset{\beta \in \mathbb{R}^{p\times B}}{\operatorname{argmin}} L_n^{PLS}(\beta) = \underset{\beta \in \mathbb{R}^{p\times B}}{\operatorname{argmin}} \sum_{s=1}^{B} \left\{ \beta(s)^\top \mathbf{H}_n(s)\beta(s) - 2\boldsymbol{h}_n(s)\beta(s) \right\}, \tag{2.6}$$

where for all $s \in \{1, \dots, B\}$, $\mathbf{H}_n(s)$ are $p \times p$ symmetrical positive semidefinite matrices and $\boldsymbol{h}_n(s)$ are p-dimensional vectors equal to

$$\mathbf{H}_n(s) = \frac{1}{n}\sum_{i=1}^{n}\int_{[0,\tau]} Y_i^s(t)\Big(X_i(t) - \bar{X}^s(t)\Big)^{\otimes 2} dt \text{ and } \boldsymbol{h}_n(s) = \frac{1}{n}\sum_{i=1}^{n}\int_{[0,\tau]} Y_i^s(t)\Big(X_i(t) - \bar{X}^s(t)\Big) dN_i(t)$$

with $\bar{X}^s(t) = \sum_{i=1}^n X_i(t)Y_i^s(t) / \sum_{i=1}^n Y_i^s(t)$ and the convention that $0/0 = 0$.

On the other hand, an estimator $\hat{\beta}_{C/add}$ in the constant coefficient model is defined as

$$\hat{\beta}_{C/add} \in \underset{\beta \in \mathbb{R}^{p}}{\operatorname{argmin}} \left(\beta^\top \mathbf{H}_n \beta - 2\boldsymbol{h}_n\beta\right), \text{ with } \mathbf{H}_n = \sum_{s=1}^{B} \mathbf{H}_n(s) \text{ and } \boldsymbol{h}_n = \sum_{s=1}^{B} \boldsymbol{h}_n(s). \tag{2.7}$$

This formula gives an analogue of the so-called stratified Cox model to the Aalen case. Note also that the quantities $\mathbf{H}_n$ and h_n are not the same as the ones involved in a standard Aalen model with no stratification (see for instance the terms D_n and d_n in Martinussen & Scheike (2009b)). In our constant coefficient models β_0 is constant and therefore its estimators are also constants, but the baseline α_0 is still stratified with respect to s.

2.4 *A total-variation penalty*

To overcome the possible over-parametrization of models (2.2) and (2.3), we propose to define penalized versions of criteria (2.4) and (2.6). For all $\beta = (\beta(s), s = 1, \dots, B)$ with $\beta(s) = (\beta^1(s), \dots, \beta^p(s))$, define for all $j = 1, \dots, p$

$$\beta^j = (\beta^j(1), \dots, \beta^j(B)) \text{ and } \mathrm{TV}(\beta^j) = \sum_{s=2}^{B} |\beta^j(s) - \beta^j(s-1)| = \sum_{s=2}^{B} |\Delta\beta^j(s)|. \tag{2.8}$$

We now consider the minimizers of the partial log-likelihood (respectively the partial least-squares) penalized with a covariate specific total variation. Define the penalized estimators in models (2.2) and (2.3) as:

$$\hat{\beta}_{\mathrm{TV}/mult} \in \underset{\beta \in \mathbb{R}^{p \times B}}{\operatorname{argmin}} \left\{ L_n^{PL}(\beta) + \frac{\lambda_n}{n} \sum_{j=1}^{p} \mathrm{TV}(\beta^j) \right\} \text{ and} \tag{2.9}$$

$$\hat{\beta}_{\mathrm{TV}/add} \in \underset{\beta \in \mathbb{R}^{p \times B}}{\operatorname{argmin}} \left\{ L_n^{PLS}(\beta) + \frac{\lambda_n}{n} \sum_{j=1}^{p} \mathrm{TV}(\beta^j) \right\}. \tag{2.10}$$

These penalized algorithms are part of the class of fused-lasso algorithms (see e.g. Tibshirani et al. (2005), Rinaldo (2009) for a definition) and can be rewritten as lasso algorithms (the details are given in Supplementary Material).

3. Asymptotic results

We successively provide the asymptotic results for the estimators $\hat{\beta}_{\mathrm{TV}/add}$ in the additive model and $\hat{\beta}_{\mathrm{TV}/mult}$ in the multiplicative model. Their proofs are postponed in Supplementary Material. In both models, the following conditions are mandatory.

Assumption 3

1. The covariate process $X(\cdot)$ is of bounded variation on $[0,\tau]$.
2. There exists a constant M such that for all t in $[0,\tau]$, $X(t) \in [-M, M]^p$ almost surely.
3. For all $s = 1, \dots, B$, $\int_{A_s^\tau} \alpha_0(t,s)dt < \infty$.

Parts 1 and 2 of Assumption 3 are equivalent to Assumption (ii) in Scheike (2002). Part 3 is a stratified form of Assumption 7.2.1, for instance, in Andersen et al. (1993).

Define, for all $s = 1, \dots, B$ and t in A_s^τ, the centered process

$$M^s(t) = N(t) - \int_0^t \mathbb{E}\big[dN(r) \mid X(r), D \wedge C \geqslant r, N(r-) = s-1\big]$$

and the $p \times p$ matrix

$$\mathbf{H}(s) := \int_{A_s^\tau} \mathbb{E}[Y^s(t) X(t)^\top X(t)]dt - \int_{A_s^\tau} \frac{(\mathbb{E}[Y^s(t)X(t)])^{\otimes 2}}{\mathbb{E}[Y^s(t)]} dt,$$

which, thanks to Assumptions 1 and 2, is well defined.

Theorem 3.1 Assume that, for each $s = 1, \dots, B$, $\mathbf{H}(s)$ is non-singular and that Assumptions 1, 2 and 3 are fulfilled.

1. If $\lambda_n/n \to 0$ as $n \to \infty$ then $\hat{\beta}_{\mathrm{TV}/add}$ converges to β_0 in probability.
2. If $\lambda_n/\sqrt{n} \to \lambda_0 \geqslant 0$ as $n \to \infty$ then $\sqrt{n}(\hat{\beta}_{\mathrm{TV}/add} - \beta_0)$ converges in distribution to

$$\underset{u\in\mathbb{R}^{p\times B}}{\operatorname{argmin}}\ \Lambda_{add}(u) = \underset{u\in\mathbb{R}^{p\times B}}{\operatorname{argmin}}\ \Big[\sum_{s=1}^{B}\big\{u(s)^\top \mathbf{H}(s)u(s) - 2u(s)^\top \xi_{add}(s)\big\}$$
$$+ \lambda_0 \sum_{j=1}^{p}\sum_{s=2}^{B}\Big\{|\Delta u^j(s)|\mathbf{1}(\Delta\beta_0^j(s) = 0) + \operatorname{sgn}(\Delta\beta_0^j(s))(\Delta u^j(s))\mathbf{1}(\Delta\beta_0^j(s) \neq 0)\Big\}\Big],$$

and for each s, $\xi_{add}(s)$ is a centered p-dimensional gaussian vector with covariance matrix equal to

$$\mathbb{E}\left[\Big(\int_{A_s^\tau}(X(t) - \mathbb{E}[Y^s(t)X(t)]/\mathbb{E}[Y^s(t)])Y^s(t)dM^s(t)\Big)^{\otimes 2}\right].$$

We now state an analogous result in the multiplicative model. Define for all $s = 1, \dots, B$ and for all t in A_s^τ,

$$s^{(l)}(s,t,\beta) = \mathbb{E}[Y^s(t)X(t)^{\otimes l}\exp(X(t)\beta(s))], l = 0, 1, 2.$$

Introduce $\mathbf{e}(s,t,\beta) = s^{(1)}(s,t,\beta)/s^{(0)}(s,t,\beta)$, $\mathbf{v}(s,t,\beta) = s^{(2)}(s,t,\beta)/s^{(0)}(s,t,\beta) - \mathbf{e}(s,t,\beta)^{\otimes 2}$ and $\mathbf{\Sigma}(s,\beta) = \int_{A_s^\tau}\mathbf{v}(s,t,\beta)\mathbb{E}[Y^s(t)dN(t)]$. For any $s = 1, \dots, B$ and for any t in A_s^τ, the three functions $s^{(l)}(s,t,\beta_0)$ are bounded due to Assumption 3 and $\mathbf{e}(s,t,\beta), \mathbf{v}(s,t,\beta)$ and $\mathbf{\Sigma}(s,\beta)$ are finite due to Assumptions 1 and 3.

Theorem 3.2 Assume that for each $s = 1, \dots, B$, $\mathbf{\Sigma}(s,\beta_0)$ is non-singular and that Assumptions 1, 2 and 3 are fulfilled.

1. If $\lambda_n/n \to 0$ as $n \to \infty$ then $\hat{\beta}_{\mathrm{TV}/mult}$ converges to β_0 in probability.
2. If $\lambda_n/\sqrt{n} \to \lambda_0 \geqslant 0$ as $n \to \infty$ then $\sqrt{n}(\hat{\beta}_{\mathrm{TV}/mult} - \beta_0)$ converges in distribution to

$$\underset{u\in\mathbb{R}^{p\times B}}{\operatorname{argmin}}\ \Lambda_{mult}(u) = \underset{u\in\mathbb{R}^{p\times B}}{\operatorname{argmin}}\ \Big[\sum_{s=1}^{B}\Big\{\frac{1}{2}u(s)^\top \mathbf{\Sigma}(s,\beta_0)u(s) - u(s)^\top \xi_{mult}(s)\Big\}$$
$$+ \lambda_0 \sum_{j=1}^{p}\sum_{s=2}^{B}\Big\{|\Delta u^j(s)|\mathbf{1}(\Delta\beta_0^j(s) = 0) + \operatorname{sgn}(\Delta\beta_0^j(s))(\Delta u^j(s))\mathbf{1}(\Delta\beta_0^j(s) \neq 0)\Big\}\Big],$$

and for each s, $\xi_{mult}(s)$ is a centered p-dimensional gaussian vector with covariance matrix equal to

$$\mathbb{E}\left[\left(\int_{A_s^\tau}\left(X(t)-\mathbf{e}(s,t,\beta_0)\right)Y^s(t)dM^s(t)\right)^{\otimes 2}\right].$$

First results of Theorems 3.1 and 3.2 prove the consistency of our estimators, on the mandatory condition that λ_n/n tends to 0. This ensures that they behave better than the constant coefficient estimators when β_0 is non constant. In addition, the considered penalty will induce sparsity for each covariate $j=1,\ldots,p$ in the successive differences $\Delta\beta^j(s)$, $s=1,\ldots,B$. As a consequence, the effects of a covariate on two consecutive events will often be equal. We show, in the following simulation study, that this induced sparsity ameliorates the behaviour of our estimators compared to the unconstrained ones (defined in Equations (2.4) and (2.6)).

The second results show that asymptotic normality can be achieved only if $\lambda_n/\sqrt{n}$ tends to 0. In that case, the asymptotic variance of the limiting distribution can be estimated by means of the analog of the optional variation in this context, see Martinussen & Scheike (2006, page 150-151) for details.

However, when $\lambda_0=0$, the algorithm is no longer consistent in selection, in the sense that in both multiplicative and additive cases, as n tends to infinity,

$$\mathbb{P}\big[\{(j,s)\in\mathbb{R}^p\times\mathbb{R}^B,\Delta(\hat{\beta}^j_{TV})(s)\neq 0\}=\{(j,s)\in\mathbb{R}^p\times\mathbb{R}^B,\Delta(\beta^j_0)(s)\neq 0\}\big]\to 0.$$

Even in the case where $\lambda_n/\sqrt{n}\to\lambda_0>0$, this probability is asymptotically stricly less than 1, see Zou (2006) for details. To enhance the sparsity in the covariate-specific successive differences (or equivalently to force $\hat{\beta}_{TV/mult}$ and $\hat{\beta}_{TV/add}$ to have several constant coefficients), we consider a reweighted lasso, in the manner of Zou (2006) or Candès et al. (2008). In both models the two steps (or reweighted lasso) estimators are defined as

$$\tilde{\beta}_{TV/mult}\in\underset{\beta\in\mathbb{R}^{p\times B}}{\operatorname{argmin}}\left\{L^{PL}(\beta)+\frac{\lambda_n}{n}\sum_{j=1}^{p}\sum_{s=2}^{B}\frac{1}{|\Delta\hat{\beta}^j_{TV/mult}(s)|+|\Delta\hat{\beta}_0|}|\Delta\beta^j(s)|\right\}\tag{3.11}$$

and

$$\tilde{\beta}_{TV/add} =\in \underset{\beta\in\mathbb{R}^{p\times B}}{\operatorname{argmin}} \left\{ L^{PLS}(\beta) + \frac{\lambda_n}{n} \sum_{j=1}^{p}\sum_{s=2}^{B} \frac{1}{|\Delta\hat{\beta}^j_{TV/add}(s)| + |\Delta\hat{\beta}_0|} |\Delta\beta^j(s)| \right\}, \tag{3.12}$$

where $\hat{\beta}_{TV/mult}$ and $\hat{\beta}_{TV/add}$ are defined in Equations (2.9) and (2.10) and in both cases:

$$|\Delta\hat{\beta}^j_0| = \min\{|\Delta\hat{\beta}^j_{TV}(s)|, j = 1,\dots,p\ , s = 1,\dots,B\ , |\Delta\hat{\beta}^j_{TV}(s)| \neq 0\}.$$

4. Implementation and simulation studies

We compare the performance of the penalized estimators (2.9) and (2.10), the constant coefficient ones (2.5) and (2.7), and the unconstrained ones (2.4) and (2.6). To mimic the bladder tumour cancer dataset studied in Section 5, we set $p = 4$ and consider $B = 5$ recurrent events for the estimation. In the multiplicative and additive models, the sample size n varies from $n = 50 = 2.5\,pB$ to $n = 1000 \simeq (pB)^{2.3}$.

4.1 *Implementation of penalized estimators.*

The minimizers of the partial log-likelihood, respectively, the partial least-squares, penalized with a covariate specific total variation of Equations (2.9) and (2.10) can be seen as lasso estimators by introducing a block matrix D of size $(pB \times pB)$ with p diagonal blocks being equal to a lower triangular matrix with nonzero elements equal to 1 and $p^2 - p$ off-diagonal blocks being matrices of zeros.

The minimization problems of Equations (2.9) and (2.10) can then be rewritten as

$$\begin{aligned}
&\hat{\beta}_{TV} = D\hat{\gamma}_{TV} \quad \text{with} \\
&\hat{\gamma}_{TV} \in \underset{\gamma\in\mathbb{R}^{p\times B}}{\operatorname{argmin}} \left\{ L_n(D\gamma) + \frac{\lambda_n}{n} \sum_{j=1}^{p}\sum_{s=2}^{B} |\gamma^j(s)| \right\} = \underset{\gamma\in\mathbb{R}^{p\times B}}{\operatorname{argmin}} \left\{ L_n(D\gamma) + \frac{\lambda_n}{n} \|\gamma\|_1 \right\},
\end{aligned} \tag{4.13}$$

where L_n is either L_n^{PL} or L_n^{PLS} and, in both cases, $\hat{\gamma}_{TV} = (\hat{\beta}^1_{TV}(1), \Delta\hat{\beta}^1_{TV}(2), \dots, \Delta\hat{\beta}^p_{TV}(B))^\top$. The related R functions can be found at `http://www.lsta.upmc.fr/guilloux.php?main=publications`.

The regularization parameter is chosen in both multiplicative and additive cases via 5-fold cross-validation, defined by the lasso formulations of Equation (4.13). For details in the multiplicative model, see Simon et al. (2011, pages 9-10) or van Houwelingen et al. (2006). Details for the additive model may be found in Martinussen & Scheike (2009b, page 608).

4.2 *Simulation scheme*

We draw $p = 4$ constant covariates from uniform distributions on $[0, 2]$ and set the parameters values at $\beta_0^1 = 0.25(0, 0, 1, 1, 0)$, $\beta_0^2 = (1, \ldots, 1)$, $\beta_0^3 = b(1, 2, 3, 4, 5)$ and $\beta_0^4 = (0, \ldots, 0)$. We generated recurrent event times from the multiplicative (2.2) and additive (2.3) models with baseline defined through the Weibull distribution with shape parameter $a_{\mathcal{W}}$ and scale parameter 1 (see Supplementary Materials for a more detailed description of the simulation scheme). The death and censoring times are generated from exponential distributions with parameters a_D and a_C respectively. We set the value of parameter $a_{\mathcal{W}}$ at 1.5, and of b at 4 in the additive case and -1 in the multiplicative case. Finally, the values of a_D and a_C are empirically determined to obtain $14-15\%$ of individuals experiencing the fifth event. More results for $\mathrm{P_{obs}} = 28-29\%$ were obtained and are reported in Supplementary Material.

4.3 *Performance evaluation*

To evaluate the performance of the different estimators, we conduct a Monte Carlo study with $M = 500$ experiences. The estimation accuracy is investigated for each method via a mean squared rescaled error defined as

$$\mathrm{MSE} = \frac{10^3}{M} \sum_{m=1}^{M} \frac{\|\hat{\beta}_m - \beta_0\|^2}{\|\beta_0\|^2}, \tag{4.14}$$

where $\hat{\beta}_m$ is the estimation in the sample m. We furthermore study the detection power of non-constant (respectively constant) covariate effects by computing specificities (SPEC) and sensitivities (SENS) for each method. For an estimation $\hat{\beta}_m$, define false (FP) and true (TP) positives

$$\text{FP}(\hat{\beta}_m) = \text{Card}\Big(j \in \{1,\dots,p\} \text{ such that } \text{TV}(\hat{\beta}_m^j) \neq 0 \text{ and } \text{TV}(\beta_0^j) = 0\Big)$$
$$\text{TP}(\hat{\beta}_m) = \text{Card}\Big(j \in \{1,\dots,p\} \text{ such that } \text{TV}(\hat{\beta}_m^j) \neq 0 \text{ and } \text{TV}(\beta_0^j) \neq 0\Big)$$

and false (FN) and true (TN) negatives by exchanging $=$ and $\neq$ in the above definitions (see (2.8) for the definition of TV). The specificity and sensibility of a method are defined over the M replications as

$$\text{SPEC} = \frac{1}{M}\sum_{m=1}^{M} \frac{\text{TN}(\hat{\beta}_m)}{\text{TN}(\hat{\beta}_m) + \text{FP}(\hat{\beta}_m)} \tag{4.15}$$

and

$$\text{SENS} = \frac{1}{M}\sum_{m=1}^{M} \frac{\text{TP}(\hat{\beta}_m)}{\text{TP}(\hat{\beta}_m) + \text{FN}(\hat{\beta}_m)}, \tag{4.16}$$

such that 1 is the ideal value for both indicators of detection power. The results are presented in Tables 1-2.

4.4 *Results*

As expected, the constant model is biased and, in particular, for our choice of a non-constant β_0, the MSE does not decrease with the sample size n. The comparison between the unconstrained and penalized estimators is in favour of our estimators in all cases as long as n is much smaller than $(pB)^2$. For $p = 4$, $B = 5$, $n = 100$ and $\text{P}_{\text{obs}} = 14\%$ (which are values close to those encountered in the bladder tumour cancer dataset studied in the next section) our penalized estimators are respectively, 2.41 in the additive model and 1.38 in the multiplicative model, times better than the unconstrained ones in terms of estimation error.

Regarding the sensitivity and specificity indexes defined in Section 4.3, the unconstrained estimators, which, by definition, has no constant coefficients, has a perfect sensitivity and a null specificity. On the opposite, the constant estimator can not detect a non-constant effect and, as a consequence, produces no false positive, nor true positive, with the consequence that its sensitivity is null and its specificity equals one.

The one-step penalized estimators of Equations (2.9) and (2.10) have almost perfect sensitivities, but their specificities decrease with the sample size. In our opinion, this is due to the cross validation, which tends to choose smaller regularization parameters for larger sample sizes. This choice is consistent with the theoretical values of λ_n displayed in Theorems 3.1 and 3.2. This however leads to the classical non-consistency in selection of one-step estimators, as discussed at the end of Section 3.

To overcome this drawback, we defined in Equations (3.11) and (3.12), two-steps estimators. They are expected to enhance the sparsity in the estimated successive differences and consequently, produce less false positives and more true negatives. Results indeed show increased specificities for comparable sensitivities, as compared to the one-step estimators. These two-steps estimators however are expected to be more biased, this phenomenon can be seen in Table 2 for example. When the specificities of the one-step and two-steps estimators are of the same order, the MSE of the latter is greater.

We repeat the simulation study for $a_{\mathcal{W}} = 0.5$ and then for a Gompertz baseline with shape parameter $a_{\mathcal{G}} = 1.5$ and 0.5 and scale parameter 1. The results are reported in Supplementary Material. As expected, we observe that a drop in the percentage of individuals experiencing the fifth event drops affects the performances of all estimators. Other conclusions are similar.

5. Bladder tumour data analysis

In this section we illustrate the behaviour of our estimators on the bladder tumour cancer data of Byar (1980). These data were obtained from a clinical trial conducted by the Veterans Administration Co-operative Urological Group. One hundred and eighteen patients were randomised to one of three treatments: placebo, pyridoxine or thiotepa. For each patient, the time of recurrence tumours were recorded until the death or censoring times. The number of recurrences ranges from

0 to 10. Two patients who were censored before the beginning of the study are removed. The dataset is therefore composed of 116 patients with 47 patients from the placebo group, 38 from the thiotepa group and 31 from the pyridoxine group. On these patients, since 13.79% experienced at least five tumour recurrences and only 6.9% patients experienced six tumour recurrences or more, we set the parameter B to 5. For interpretation purpose, the treatment variable is coded as two new binary variables, pyridoxine and thiotepa, making placebo the reference. In addition to these two treatment variables two supplementary covariates were recorded for each patient: the number of initial tumours and the size of the largest initial tumour.

Tables 3, 4 and figure 1 display the estimates obtained from the constant coefficient, unconstrained, total variation and two steps total variation estimators in the multiplicative model. The unconstrained estimator shows very strong variations and is difficult to interpret as such. On the other hand, the constant coefficient estimator gives valuable information on the impact of each covariate, but in turn cannot detect a change in variation. Our total-variation estimators reach compromise: they are not constant but easily interpretable.

For instance, a remarkable aspect of the pyrodixine treatment can be highlighted from the total variation estimation: this treatment produces a protective effect for the first three tumour recurrences but the hazard rate of further recurrences are increased by this treatment. In the same way, an increase in the effect of the initial number of tumours on recurrences is observed from the third recurrence. On the opposite, the effects of the thiotepa treatment or the size of the largest tumour are shown to be constant in the total variation model, the parameter estimates having values similar to the ones obtained in the constant model.

Our conclusions on the treatments effects are in agreement with previous studies on bladder tumours recurrences. For instance, no difference in the rate or time to tumour recurrence was found from patients using pyrodixine with patients using placebo in Tanaka et al. (2011) and Goossens et al. (2012). Moreover, Huang & Chen (2003) and Sun et al. (2006) have respectively studied gap time recurrences in the multiplicative and additive models. The results obtained from the former showed a small protective effect of this treatment while the latter concluded that gap times did not seem related to pyridoxine. These examples illustrate the nice features of our total-variation estimator: it provides sharper results, giving relevant informations on covariates effect with respect to the number of recurrent events experienced by a subject and it provides the ability to detect a change of variation. Further details are provided in Supplementary Material.

6. Discussion

In this paper, the Aalen and Cox models were studied to model the effect of covariates on the rate function. However, such models are not essential in our approach. Penalized algorithms could be easily derived for other models such as the accelerated failure time model or the semiparametric transformation model for instance.

Although we have only presented asymptotic theoretical results, the simulation studies show clear evidence that our estimators outperform standard estimators for small sample sizes. Therefore, it would be of great interest to study their finite sample properties. However, such results involve deviation inequalities for non i.i.d. and non martingale empirical processes. To our knowledge, no such results have yet been established in the context of recurrent events.

Another development of the present paper would be to establish results for the estimation of change-point locations and the number of change-points. Such results can be found for the change-point detection in the mean of a gaussian signal in Harchaoui & Lévy-Leduc (2010), for instance.

Acknowledgement

The authors are very grateful to the associate editor and the referees for their valuable comments which have helped improve the original manuscript.

References

Aalen, O. (1980). A model for nonparametric regression analysis of counting processes. In *Mathematical statistics and probability theory (Proc. Sixth Internat. Conf., Wisła, 1978)*, vol. 2 of *Lecture Notes in Statist.* New York: Springer, pp. 1–25.

Andersen, P. K., Borgan, Ø., Gill, R. D. & Keiding, N. (1993). *Statistical models based on counting processes.* Springer Series in Statistics. New York: Springer-Verlag.

Andersen, P. K. & Gill, R. D. (1982). Cox's regression model for counting processes: a large sample study. *Ann. Statist.* **10**, 1100–1120.

Bouaziz, O., Comte, F. & Guilloux, A. (2013). Nonparametric estimation of the intensity function of a recurrent event process. *Statist. Sinica* **23**, 635–665.

Byar, D. P. (1980). The veterans administration study of chemoprophylaxis for recurrent stage 1 bladder tumors: comparison of placebo, pyridoxine, and topical thiotepa. *Bladder Tumors and Others Topics in Urological Oncology* , (Edited by M. Pavone–Macaluso, P. H. Smith, and F. Edsmyn), 363–370.

Cai, J. & Schaubel, D. E. (2004). Marginal means/rates models for multiple type recurrent event data. *Lifetime Data Anal.* **10**, 121–138.

Candès, E. J., Wakin, M. B. & Boyd, S. P. (2008). Enhancing sparsity by reweighted l_1 minimization. *J. Fourier Anal. Appl.* **14**, 877–905.

Dauxois, J.-Y. & Sencey, S. (2009). Non-parametric tests for recurrent events under competing risks. *Scand. J. Statist.* **36**, 649–670.

Gaiffas, S. & Guilloux, A. (2012). High-dimensional additive hazards models and the lasso. *Electron. J. Statist.* **6**, 522–546.

Goossens, M. E., Buntinx, F., Joniau, S., Ackaert, K., Ameye, F., Billiet, I., Braeckman, J., Breugelmans, A., Darras, J., Dilen, K., Goeman, L., Kellen, E., Tombal, B., Van Bruwaene, S., Van Cleyenbreuge, B., Van der Aa, F., Vekemans, K., Van Poppel, H. & Zeegers, M. (2012). Designing the selenium and bladder cancer trial (seleblat), a phase lll randomized chemoprevention study with selenium on recurrence of bladder cancer in belgium. *BMC Urology* **12**.

Harchaoui, Z. & Lévy-Leduc, C. (2010). Multiple change-point estimation with a total variation penalty. *J. Amer. Statist. Assoc.* **105**, 1480–1493.

Huang, J., Sun, T., Ying, Z., Yu, Y. & C.-H., Z. (2013). Oracle inequalities for the lasso in the cox model. *Ann. Statist.* **41**, 1142–1165.

Huang, Y. & Chen, Y. Q. (2003). Marginal regression of gaps between recurrent events. *Lifetime Data Anal.* **9**, 293–303.

Huffer, F. W. & McKeague, I. (1991). Weighted least squares estimation for aalen's additive risk model. *J. Amer. Statist. Assoc.* **86**, 114–129.

Kalbfleisch, J. D. & Prentice, R. L. (2011). *The statistical analysis of failure time data*, vol. 360. John Wiley & Sons.

Kelly, P. & Lim, L. L.-Y. (2000). Survival analysis for recurrent event data: an application to childhood infectious diseases. *Statist. Med.* **19**, 13–33.

Lawless, J. F. & Nadeau, C. (1995). Some simple robust methods for the analysis of recurrent events. *Technometrics* **37**, 158–168.

Lin, D. Y., Wei, L. J., Yang, I. & Ying, Z. (2000). Semiparametric regression for the mean and rate functions of recurrent events. *J. R. Statist. Soc. Ser. B Statist. Methodol.* **62**, 711–730.

Lin, D. Y., Wei, L. J. & Ying, Z. (1998). Accelerated failure time models for counting processes. *Biometrika* **85**, 605–618.

Lin, D. Y. & Ying, Z. (1994). Semiparametric analysis of the additive risk model. *Biometrika* **81**, 61–71.

Martinussen, T. & Scheike, T. H. (2006). *Dynamic regression models for survival data.* Statistics for Biology and Health. New York: Springer.

Martinussen, T. & Scheike, T. H. (2009a). The additive hazards model with high-dimensional regressors. *Lifetime Data Anal.* **15**, 330–342.

Martinussen, T. & Scheike, T. H. (2009b). Covariate selection for the semiparametric additive risk model. *Scand. J. Statist.* **36**, 602–619.

McKeague, I. W. (1988). Asymptotic theory for weighted least squares estimators in Aalen's additive risk model. In *Statistical inference from stochastic processes (Ithaca, NY, 1987)*, vol. 80 of *Contemp. Math.* Providence, RI: Amer. Math. Soc., pp. 139–152.

Prentice, R. L., Williams, B. J. & Peterson, A. V. (1981). On the regression analysis of multivariate failure time data. *Biometrika* **68**, 373–379.

Rinaldo, A. (2009). Properties and refinements of the fused lasso. *Ann. Statist.* **37**, 2922–2952.

Scheike, T. H. (2002). The additive nonparametric and semiparametric Aalen model as the rate function for a counting process. *Lifetime Data Anal.* **8**, 247–262.

Simon, N., Friedman, J., Hastie, T. & Tibshirani, R. (2011). Regularization paths for cox's proportional hazards model via coordinate descent. *J. Statist. Software* **39**, 1–13.

Sun, L., Park, D.-H. & Sun, J. (2006). The additive hazards model for recurrent gap times. *Statist. Sinica* **16**, 919–932.

Tanaka, T., Miyazawa, K., Tsukamoto, T., Kuno, T. & Suzuki, K. (2011). Pathobiology and chemoprevention of bladder cancer. *J. Oncol.* .

Therneau, T. M. (2000). *Modeling survival data: extending the Cox model*. Springer.

Tibshirani, R. (1997). The lasso method for variable selection in the cox model. *Statist. Med.* **16**, 385–395.

Tibshirani, R., Saunders, M., Rosset, S., Zhu, J. & Knight, K. (2005). Sparsity and smoothness via the fused lasso. *J. R. Statist. Soc. Ser. B Statist. Methodol.* **67**, 91–108.

van Houwelingen, H., Bruinsma, T., Hart, A., van't Veer, L. & L.F.A., W. (2006). Cross- validated cox regression on microarray gene expression data. *Statist. Med.* **25**, 3201–3216.

Wei, L. J., Lin, D. Y. & Weissfeld, L. (1989). Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. *J. Amer. Statist. Assoc.* **84**, 1065–1073.

Zeng, D. & Cai, J. (2010). A semiparametric additive rate model for recurrent events with an informative terminal event. *Biometrika* **97**, 699–712.

Zou, H. (2006). The adaptive lasso and its oracle properties. *J. Amer. Statist. Assoc.* **101**, 1418–1429.

Supplementary material

Supplementary material contains proofs of Theorems 3.1 and 3.2. They also include comments on the asymptotic distribution of our estimators, details of the simulation scheme used in Section 4, an extended simulation study and additional analysis on the bladder tumour data of Byar (1980).

Tables

Table 1. *Simulation results in the multiplicative model for* $P_{obs} = 14\%$

n	Unconstrained			Constant			TV			two-steps TV		
	MSE	SPEC	SENS	MSE	SPEC	SENS	MSE	SPEC	SENS	MSE	SPEC	SENS
50	5576.511	0	1	225.077	1	0	72.732	0.271	0.813	67.697	0.598	0.68
100	64.231	0	1	216.658	1	0	46.484	0.226	0.844	39.562	0.583	0.709
500	12.447	0	1	212.578	1	0	17.232	0.18	0.911	17.998	0.917	0.559
1000	9.06	0	1	213.292	1	0	14.215	0.192	0.9	17.353	0.983	0.512

MSE: mean square error, SPEC: specificity, SENS: sensitivity.

Table 2. *Simulation results in the additive model for* $P_{obs} = 14\%$

n	Unconstrained			Constant			TV			two-steps TV		
	MSE	SPEC	SENS	MSE	SPEC	SENS	MSE	SPEC	SENS	MSE	SPEC	SENS
50	1208.174	0	1	398.849	1	0	367.377	0.312	1	480.105	0.601	0.992
100	534.269	0	1	360.757	1	0	221.454	0.241	1	283.258	0.582	1
500	202.669	0	1	339.446	1	0	139.481	0.154	1	171.794	0.525	1
1000	168.751	0	1	337.813	1	0	133.39	0.103	1	157.899	0.471	1

MSE: mean square error, SPEC: specificity, SENS: sensitivity.

Table 3. *Unconstrained and constant parameters estimates for the bladder data in the multiplicative model*

s	Unconstrained				Constant			
	PYRIDOXINE	THIOTEPA	SIZE	NUMBER	PYRIDOXINE	THIOTEPA	SIZE	NUMBER
1	-0.497	-0.711	-0.028	0.202	-0.037	-0.374	0.03	0.155
2	0.466	0.013	0.044	0.014	-0.037	-0.374	0.03	0.155
3	-0.211	0.027	0.129	0.250	-0.037	-0.374	0.03	0.155
4	0.717	-0.095	0.064	0.274	-0.037	-0.374	0.03	0.155
5	0.657	-0.283	0.072	0.198	-0.037	-0.374	0.03	0.155

Table 4. *Total variation and two-steps total variation parameters estimates for the bladder data in the multiplicative model*

s	TV				two-steps TV			
	PYRIDOXINE	THIOTEPA	SIZE	NUMBER	PYRIDOXINE	THIOTEPA	SIZE	NUMBER
1	-0.080	-0.373	0.023	0.133	-0.167	-0.387	0.031	0.122
2	-0.080	-0.373	0.023	0.133	-0.167	-0.387	0.031	0.122
3	-0.080	-0.373	0.023	0.133	-0.167	-0.387	0.031	0.265
4	0.241	-0.373	0.066	0.241	0.625	-0.387	0.031	0.265
5	0.241	-0.373	0.066	0.241	0.625	-0.387	0.031	0.265

Figure Captions

Fig. 1. Estimates for the bladder data in the multiplicative model. The crosses represent the constant estimator, the filled circles the unconstrained estimator, the circles the total variation estimator and the squares the two steps total variation estimator.

Figures

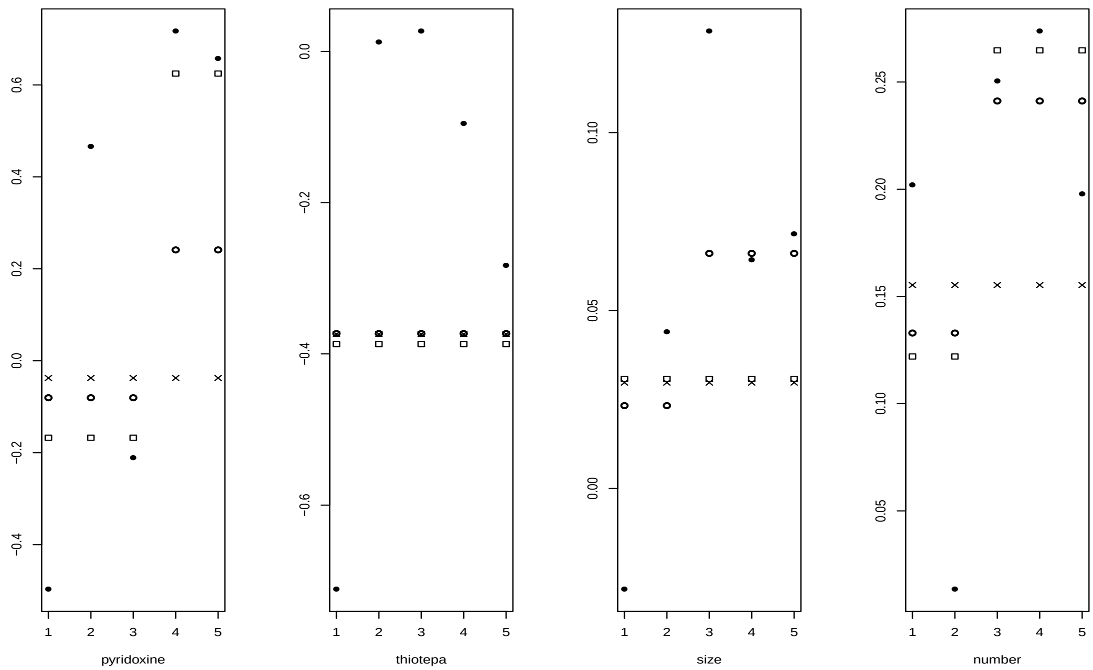

Fig. 1. Estimates for the bladder data in the multiplicative model. The crosses represent the constant estimator, the filled circles the unconstrained estimator, the circles the total variation estimator and the squares the two steps total variation estimator.